Détection et quantification de la pollution dans le réseau P2P KAD

Thibault Cholez, Guillaume Montassier, Guillaume Doyen, Rida Khatoun, Isabelle Chrisment, Olivier Festor

Groupement d'Intéret Scientifique
Surveillance, Sureté et Sécurité des Grands Systèmes
– GIS 3SGS –

Projet ACDA-P2P :
Approche Collaborative pour la Détection d'Attaques
dans les réseaux Pair à Pair

Délivrable 3

Détection et quantification de la pollution dans le réseau P2P KAD

UMR STMR 6279 :
Thibault Cholez
Guillaume Montassier
Guillaume Doyen
Rida Khatoun

LORIA–INRIA Nancy Grand Est :
Isabelle Chrisment
Olivier Festor

21 octobre 2011

Résumé

Ce délivrable présente les résultats des travaux menés durant le troisième trimestre (T0+9) du projet GIS 3SGS Acda-P2P dont l'objectif est de proposer une architecture collaborative pour la détection d'attaques dans les réseaux pair à pair. Nous décrivons dans ce rapport nos travaux concernant la détection et la quantification de la pollution dans le réseau P2P KAD (tâche T3). Ces travaux [1] ont été publiés dans conférence internationale de référence concernant la recherche sur les réseaux P2P, à savoir : IEEE P2P 2011 [MCD+11]. Ils seront étendus dans le dernier délivrable qui présentera la mise en oeuvre de la détection de la pollution de manière collaborative (T5).

Nous introduisons tout d'abord ce rapport en situant nos travaux courants par rapport à la problématique de la pollution dans les réseaux P2P et proposons une classification des quelques études précédentes réalisées dans ce domaine.

Nous décrivons ensuite une forme particulière de pollution affectant le réseau KAD et présentons notre solution capable d'évaluer la pollution d'un fichier partagé au sein de ce réseau. Cette première contribution théorique consiste d'une part en l'analyse de l'architecture de KAD afin de permettre l'obention des informations révélant la pollution, et d'autre part, en la conception d'une métrique capable de d'appréhender ces informations afin d'y détecter la pollution.

Dans un second temps, nous procédons à la validation expérimentale de notre métrique et à la quantification de la pollution. Nous collectons ainsi les informations concernant 2000 fichiers populaires au sein du réseau et utilisons un échantillon de ces données afin de confronter notre métrique à l'avis d'experts. Une fois notre métrique validée, nous quantifions et caractérisons la pollution affectant le réseau KAD. Nous montrons que la grande majorité des fichiers populaires sont pollués et pointent vers des contenus indésirables, notamment pornographiques.

Nous terminons ce rapport par une conclusion et l'annonce des derniers travaux envisagés dans le cadre du projet GIS 3SGS Acda-P2P.

1. http://hal.inria.fr/inria-00619965/PDF/P2P11-KAD_pollution_quantification-Cholez.pdf

Table des matières

Chapitre 1

Introduction

Le projet Acda-P2P est un projet académique, financé par le GIS [1] 3SGS [2]. Il regroupe l'équipe ERA [3] de l'UMR 6279 STMR [4] et l'équipe MADYNES [5] de l'INRIA [6] Grand Est. D'une durée initiale d'un an, le projet a débuté en mai 2010. Il vise à proposer une solution collaborative pour la détection d'attaques dans les réseaux pair à pair (P2P). Plus spécifiquement, il s'inscrit dans le contexte scientifique suivant, détaillé dans le premier délivrable et résumé ici.

Les réseaux pair à pair (P2P), notamment ceux utilisant les tables de hachage distribuées, sont devenus, en quelques années, une application majeure de l'Internet en permettant à des millions d'utilisateurs de partager rapidement et sans coût d'infrastructure de grandes quantités de données. Cependant, les réseaux P2P peuvent également être un support pour des activités malveillantes menaçant la sécurité du réseau P2P lui-même (pollution des données, surveillance des échanges, ...) ou, plus généralement, d'Internet (déni de service, propagation de vers, contrôle de botnet, ...). Étant donné le développement croissant de ces réseaux, pouvoir détecter lorsqu'un réseau P2P devient le support d'activités malveillantes devient primordial pour s'en prémunir.

Le premier délivrable a présenté les différentes architectures P2P parmi lesquelles les Table de Hachage Distribuées (DHT), qui ont été retenues pour notre étude de part leurs qualités intrinsèques et leur déploiement à grande échelle. Celles-ci souffrent néanmoins de nombreuses attaques que nous avons précédemment décrites. En particulier, la vulnérabilité la plus critique consiste en l'insertion ciblée de nœuds malveillants pouvant prendre le contrôle des références stockées au sein du réseau et que nous avons appelé attaque interne localisée dans la taxonomie proposée en introduction du second délivrable. Dans ce même délivrable, nous avons présenté une étude montrant que la DHT utilisée par BitTorrent pour distribuer les trackers est notamment vulnérable à cette attaque. Nous avons également proposé une manière de les détecter que nous avons appliquer à KAD et ainsi montré que le réseau était largement affecté par les attaques internes localisées. Malheureusment, une seconde collecte de données a montré que les attaques étaient variables dans le temps ce qui empêche une étude approfondie de celles-ci à

1. Groupement d'Intérêt Scientifique
2. Surveillance, Sureté et Sécurité des Grands Systèmes
3. Environnement de Réseaux Autonomes
4. Science et Technologie pour la Maîtrise des Risques
5. Management of Dynamic Networks and Services
6. Institut National de Recherche en Informatique et Automatique

court terme.

Dans ce contexte, ce délivrable présente les résultats de la tâche T4 du projet ACDA-P2P qui a été recentrée sur l'étude d'un autre problème de sécurité majeur des réseaux P2P, à savoir la pollution des contenus. Nous proposons tout d'abord une taxonomie des différentes formes de pollution permettant de mettre en évidence l'originalité de celle que nous avons identifiée sur KAD. Nous présentons ensuite notre approche permettant de détecter les fichiers pollués et la mettons en oeuvre afin de quantifier la pollution à l'échelle du réseau.

Les contributions sont organisées comme suit : le chapitre 2 rappelle le contexte de la pollution dans les réseaux P2P. Le chapitre 3 présente notre métrique permettant de détecter la falsification de l'indexation des fichiers. Nous quantifions ensuite, dans le chapitre 4, dans quelle mesure celle-ci affecte les contenus populaires partagés dans KAD. Enfin, le chapitre 5 conclut et indique les derniers travaux menés dans le projet ACDA-P2P (T5).

Chapitre 2

Contexte

Nous avons décrit et détecté, dans le précédent délivrable, des attaques reposant sur l'insertion ciblée de nœuds au sein de KAD malgré les plus récentes protections. Les nœuds ainsi placés peuvent réaliser plusieurs actions malveillantes (surveillance, attaque éclipse, déni de service, etc.) et nous avions détecté de nombreux placements suspects durant certaines périodes de mesure indiquant que le réseau est ponctuellement la cible d'attaques.

Cependant, les attaques internes nécessitannt l'insertion de noeuds ne sont pas les seuls problèmes de sécurité affectant les résaux P2P. De précédentes études [LKXR05] [LNR06] ont en effet révélé que les réseaux P2P sont largement victimes de pollution. Cependant, ces études datent de 2005 et ont été réalisées sur des réseaux cibles aujourd'hui obsolètes tels qu'Overnet ou Kazaa. Dès lors, on peut légitimement se demander si la quantification de la pollution et les différentes formes relevées alors correspondent aux réseaux actuelement opérationnels tels que KAD.

Notre objectif est ici d'étudier la pollution affectant le mécanisme d'indexation de KAD et pouvant amener les utilisateurs à télécharger inconsciemment des contenus indésirables portant atteinte à leur sécurité. Il est important de d'évaluer quels sont les types de pollution actuels ainsi que leur niveau de propagation afin de savoir comment, et dans quelle mesure, les dizaines de millions d'utilisateurs utilisant les tables de hachage distribuées publiques sont affectés par cette menace.

Les différents types de pollution

La pollution des réseaux P2P étant un sujet vaste et complexe, nous proposons tout d'abord de recenser les différentes formes de pollution pouvant s'appliquer au réseau KAD ainsi que leurs moyens possibles de mise en oeuvre afin de bien positionner notre contribution dans ce domaine. Nous distinguons deux types de pollution affectant les réseaux P2P selon qu'ils impliquent directement la corruption du contenu (*content pollution*) ou celle du système d'indexation de la DHT (*metadata pollution*), ce qui est illustré par la figure 2.1.

Pollution des contenus

La première forme de pollution par des contenus consiste à partager des fichiers dont le contenu correspond à la description mais dont celui-ci est fortement dégradé. Cette forme de

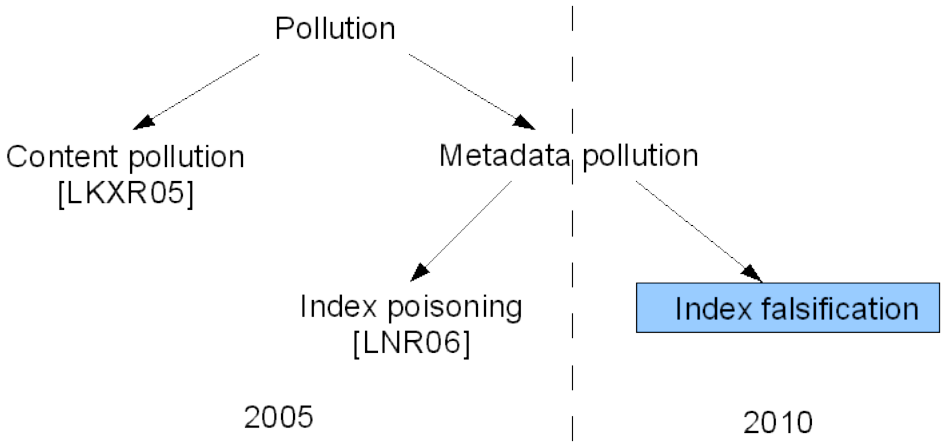

FIGURE 2.1 – Taxonomie des différentes formes de pollution

pollution a notamment été largement constatée sur le réseau Kazaa [LKXR05]. La dégradation peut avoir plusieurs formes telles que l'ajout de bruit, la fin prématurée du contenu, l'annonce d'un message, etc. Une autre forme de pollution des contenus vise à dégrader les performances de téléchargement d'un fichier existant non pollué en partageant de fausses pièces d'information. Cette forme de pollution est notamment utilisée dans le cadre de BitTorrent dont l'indexation des fichiers partagés est contrôlée par des site web qui limitent l'indexation de fichiers indésirables.

Du point de vue du pollueur, ces deux formes de pollution ont le désavantage de requérir, au moins dans un premier temps, des ressources pour diffuser rapidement les fichiers corrompus. Dans le cadre des réseaux P2P basés sur des DHTs, l'indexation des contenus partagés est réalisée par les pairs eux-mêmes et est peu contrôlée. Cette faiblesse permet une autre forme plus efficace de pollution appelée pollution de l'indexation que nous allons étudier plus en détail dans KAD.

Pollution de l'indexation

Ainsi, la pollution dite par empoisonnement d'index consiste à remplir le système d'indexation de fausses références de fichiers, qui ne sont en réalité partagés par aucune source. Cette forme de pollution a été constatée dans le réseau Overnet [LNR06] qui opérait une DHT similaire à KAD. Afin de sembler populaire, chaque faux fichier est annoncé de nombreuses fois. L'article [LMSW10] a montré que cette forme de pollution peut également affecter KAD. Si le nombre de faux fichiers annoncés pour un mot-clé donné est suffisamment élevé, ceux-ci peuvent de plus saturer la table d'indexation des pairs chargés dudit mot-clé et empêcher ainsi le référencement de fichiers légitimes.

Cependant, une autre forme de corruption existe et n'a pas été étudiée jusqu'alors. Nous l'appelons falsification d'indexation (ou mélange d'indexation). Cette pollution consiste à indexer un même fichier sous différents noms, et par conséquent différents mots-clés, qui ne sont pas liés au contenu réel du fichier. Pour un titre donné, le fichier pollué est publié de nombreuses fois par de fausses sources afin de le rendre populaire. Cette forme de pollution est

plus néfaste que les précédentes pour deux raisons. D'abord, car elle aboutit au téléchargement complet d'un contenu indésirable et gaspille par conséquent inutilement des ressources, mais surtout, car le contenu téléchargé peut être dangereux pour l'utilisateur. Le contenu réel peut ainsi être un virus ou une vidéo pouvant gravement heurter la sensibilité de l'utilisateur (contenu pornographique, pédophile, etc.). Par ailleurs, cette forme de pollution génère des faux positifs lors de la supervision des fichiers illégaux dont l'indexation est ainsi falsifiée. Or, si les utilisateurs peuvent expérimenter régulièrement cette forme de pollution sur un réseau P2P tel que KAD, aucune étude à ce jour n'a décrit ni quantifié cette forme de pollution.

Nous nous intéressons ici à la détection et à la quantification de cette dernière forme de pollution falsifiant l'indexation des fichiers car celle-ci demeure actuellement méconnue et peut fortement dégrader, outre les performances du réseau, sa sécurité et celle de ses utilisateurs.

Chapitre 3

Détection de la pollution

3.1 Visibilité de la falsification d'indexation

Le principe de cette pollution repose sur le fait d'associer de nombreux noms différents à un même fichier. Cependant, le schéma d'indexation à deux niveaux de KAD rend très difficile la détection de cette pollution. En effet, si l'on se place du point de vue des pairs responsables d'un mot-clé, ces derniers ne reçoivent que les publications, pour un fichier pollué donné, incluant le mot-clé dont ils ont la charge dans leur nom. Cette contrainte forte sur la présence d'un mot-clé empêche de détecter la falsification d'indexation au niveau des mots-clés. Dans le cas où deux noms complètement différents sont associés à un fichier, ce dernier sera indexé au travers des mots-clés sur des pairs complètement différents. L'interrogation des pairs responsables des mots-clés ne permet donc pas détecter la falsification d'indexation. Nous avons confirmé cette hypothèse en plaçant une sonde à proximité du mot-clé « avatar ». L'ensemble des noms de fichiers recueillis par la sonde contiennent effectivement le mot-clé « avatar » et semblent donc tous effectivement proposer le bon contenu alors que ce mot-clé est en réalité fortement affecté par la pollution.

Si l'on se place du point de vue des pairs responsables d'indexer les sources d'un fichier, ceux-ci ne sont pas non plus capables d'appréhender la falsification de l'indexation. En effet, bien que l'ensemble des sources publie le fichier pollué sur les quelques pairs proches de son identifiant, le nom du fichier partagé ne fait pas partie des informations publiées lors de l'envoi d'une requête de type `KADEMLIA2_PUBLISH_SOURCE_REQ` associant une source (adresse IP, port) à un fichier (identifiant MD4). Le nom du fichier partagé est une information uniquement associée aux requêtes de type `KADEMLIA2_PUBLISH_KEY_REQ` afin de guider le choix de l'utilisateur lors de la présentation des différents fichiers disponibles pour un mot-clé. L'interrogation de la DHT de KAD ne permet donc pas de détecter la falsification d'indexation. Les pairs indexant les mots-clés ne voient pas les noms de fichiers contradictoires et les pairs responsables des sources n'ont pas connaissance des noms de fichiers associés.

Il est cependant possible de détecter la falsification d'indexation grâce à une fonctionnalité des clients KAD qui est externe à la DHT et disponible lors du téléchargement d'un fichier. Lorsqu'un fichier doit être téléchargé, les sources potentielles sont découvertes par interrogation de la DHT (`KADEMLIA2_SEARCH_SOURCE_REQ`). L'utilisation de la DHT de KAD s'arrête alors et une connexion TCP vers chacune des sources potentielles est alors initiée afin de commencer le téléchargement. Les pairs acceptant la connexion constituent les sources réelles (disponibles à

un moment donné) et, selon leur charge, partagent une partie du fichier ou placent la demande dans une file d'attente. Une requête spécifique peut alors être envoyée aux sources réelles par l'intermédiaire de la connexion TCP afin de connaître le nom par lequel ces dernières partagent le fichier demandé. Cette information est accessible dans la fenêtre « Détails du fichier » de l'interface graphique. Grâce à ces informations, des noms contradictoires annoncés par les différentes sources d'un même fichier peuvent apparaître et la falsification d'indexation peut ainsi être constaté. La capture d'écran 3.1 montre ainsi les noms annoncés pour un fichier sain alors que la capture 3.2 montre un fichier dont l'indexation est corrompue. Concernant le fichier sain, nous pouvons constater que la majorité des sources (22) annoncent exactement le nom de fichier requis par l'utilisateur et les autres sources annoncent une variante minime du même nom en gardant de nombreux mots-clés en commun. En revanche, concernant le fichier pollué, aucun des pairs n'annonce le nom de fichier désiré par l'utilisateur, à savoir « Indiana Jones et les Aventuriers de l'Arche Perdue » , ni même ne s'accorde sur un autre nom, tous les noms de fichier annoncés étant différents les uns des autres et ne partageant aucun mot-clé commun.

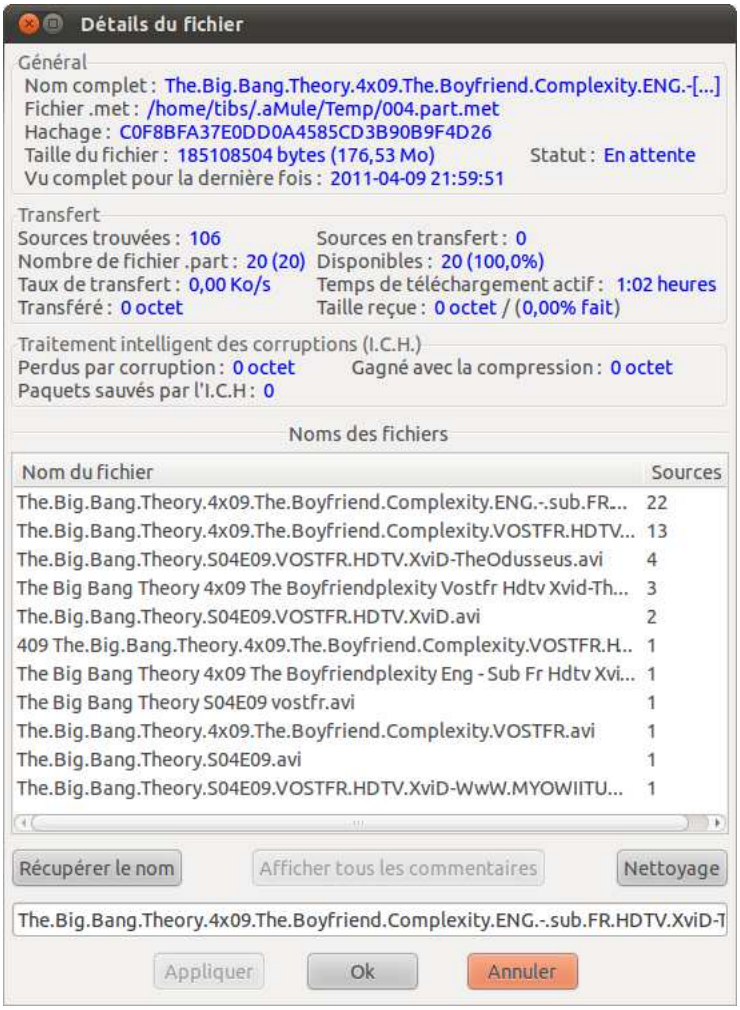

FIGURE 3.1 – Noms de fichier annoncés par les sources d'un fichier sain

FIGURE 3.2 – Noms de fichier annoncés par les sources d'un fichier pollué

3.2 Métrique de détection

Étant donné un fichier dont le contenu est identifié par son empreinte MD4, nous souhaitons savoir si celui-ci est fiable ou fait l'objet d'une pollution par falsification d'indexation en analysant les différents noms de fichier donnés par les sources contactées. Nous utilisons pour cela une métrique capable d'apprécier la similarité entre deux ensembles de mots comme la distance de Jaccard [MS99]. Soit X et Y deux ensembles de mots-clés, X étant l'ensemble des mots composant le nom de fichier choisi par l'utilisateur et Y étant l'ensemble des mots composant un des noms de fichier donné par les sources, l'indice de similarité de Tversky [Tve77] $S(X, Y)$ est une valeur entre 0 et 1 définie par la formule suivante :

$$S(X,Y) = \frac{|X \cap Y|}{|X \cap Y| + \alpha * |X - Y| + \beta * |Y - X|} \tag{3.1}$$

Pour détecter la pollution, nous utilisons un cas particulier de cette formule en fixant $\alpha = \beta = 0.5$ car aucun des deux noms ne peut être privilégié et considéré comme une référence, ce qui produit le coefficient de similarité de Dice [MS99] définit par :

$$S(X,Y) = \frac{|X \cap Y|}{|X \cap Y| + 0,5 * |X - Y| + 0,5 * |Y - X|} = \frac{2 * |X \cap Y|}{|X| + |Y|} \tag{3.2}$$

Si les deux noms de fichiers sont identiques, le coefficient résultant est de 1, si les deux noms sont complètement disjoints, le coefficient résultant vaut 0. Pour attribuer un indice de pollution P à un fichier X, nous calculons la moyenne de l'ensemble des indices de similarité obtenus par les différents noms de fichiers trouvés, soit :

$$P(X) = \frac{\sum_{i=1}^{n} S(X, Y_i)}{n} \tag{3.3}$$

Chapitre 4

Quantification et caractérisation de la pollution

4.1 Exploration des fichiers

Afin de quantifier la pollution affectant le réseau P2P KAD, nous avons collecté les informations présentées ci-avant pour de nombreux fichiers. L'accès aux noms de fichiers contradictoires est très couteux puisqu'il nécessite : (1) une recherche de fichiers sur la DHT, (2) une recherche de sources potentielles sur la DHT, (3) l'établissement d'une connexion TCP pour chaque source réelle. Nous limitons pour cela notre collecte d'information à l'étude des 100 contenus les plus téléchargés de 2010 selon l'un des principaux site d'indexation de torrents [1] recevant plus de 100 millions de recherches par an [2]. La liste des 100 mots-clés recherchés est disponible en annexe A. Pour chacun des contenus, nous collectons les différents noms associés aux 20 fichiers les plus populaires trouvés dans KAD, c'est à dire dont le nombre de sources estimé à l'issue de la recherche par mots-clés est le plus important. Nous estimons donc la pollution du réseau à partir d'une base de 2000 fichiers parmi les plus populaires.

Une fois le téléchargement d'un fichier lancé, la découverte des sources réelles est progressive et prend un certain temps que nous souhaitons estimer avant de lancer la collecte des différents noms pour les 2000 fichiers. Sur un échantillon de 150 fichiers, nous avons ainsi mesuré l'évolution du nombre de sources réelles obtenues pendant une heure dont le graphique 4.1 illustre les dix premières minutes. Il apparaît qu'en moyenne plus de 97% des sources trouvées à l'issue d'une heure sont obtenues dès 300 secondes. Pour collecter les données nécessaires à la quantification de la pollution, nous instrumentons un client KAD recherchant séquentiellement les mots-clés définis en annexe A, lançant pour chacun d'eux le téléchargement des 20 fichiers les plus populaires et enregistrant, après 300 secondes d'attente, les différents noms de fichiers obtenus par interrogation des sources réelles trouvées.

1. `http://www.kickasstorrents.com/`
2. `http://torrentfreak.com/bittorrent-zeitgeist-what-people-searched-for-in-2010-101227/`

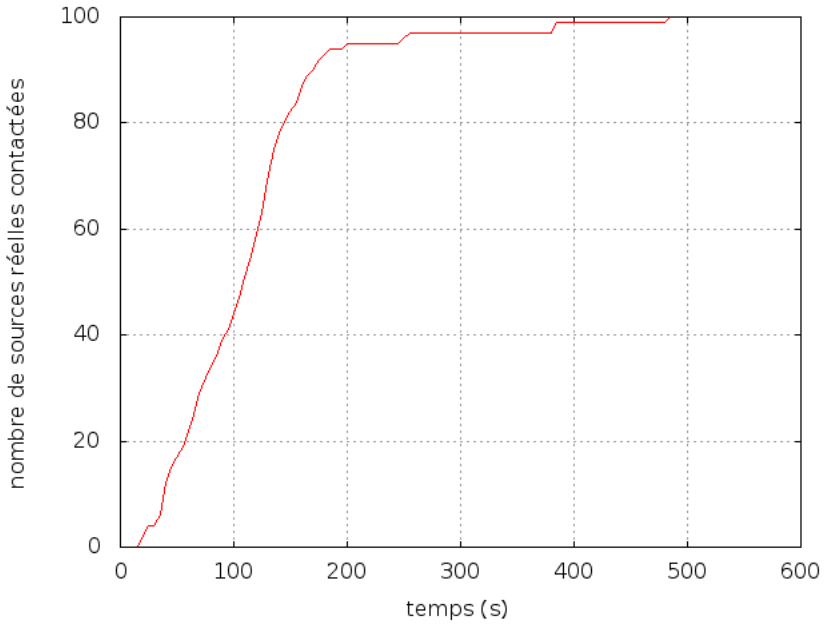

FIGURE 4.1 – Nombre moyen de sources réelles trouvées en fonction du temps

4.2 Résultats

Après avoir collecté les différents noms possibles des 2000 fichiers, nous avons appliqué notre métrique à chacun d'eux. Les graphiques 4.2 et 4.3 montrent respectivement la distribution et la distribution cumulative du nombre de fichiers en fonction des indices de pollution obtenus. Nous pouvons remarquer que la métrique est bien discriminante en établissant la majorité des scores aux extrémités de l'échelle ce qui facilite l'établissement de seuils de détection. Nous définissons ainsi trois seuils. Les fichiers dont le score de similarité est supérieur à 0.7 sont considérés comme sains, entre 0.7 et 0.3 comme peut-être pollués et inférieur à 0.3 comme pollués.

L'application de ces trois seuils à l'ensemble des 2000 fichiers populaires étudiés donne la répartition présentée dans le diagramme 4.4. Plus de 41% des contenus populaires sont ainsi clairement pollués par la falsification d'indexation. De plus, pour 21% des fichiers, aucune source réelle n'a pu être trouvée malgré le fait que les fichiers apparaissent avec un nombre élevé de sources estimées lors de la recherche par mot-clé, ceci correspond donc à une pollution par empoisonnement d'index. En considérant les deux formes de pollution, plus de 62% des fichiers sont pollués, bien que le la falsification d'indexation soit la plus néfaste des deux formes de pollution constatées. Seuls 29% des fichiers peuvent être considérés comme parfaitement sains ; le doute subsistant par rapport à la métrique pour moins de 10% des fichiers.

Afin d'évaluer la précision de notre métrique, nous avons procédé à l'analyse des noms de fichiers obtenus par dix experts. Chacun d'eux a évalué 100 fichiers dont des sources réelles ont été trouvées, soit 10% de l'échantillon global et ont classé, tout comme la métrique, chaque fichier parmi les trois catégories : pollué, peut-être pollué ou sain. Les experts ont cependant très peu recours à la catégorie « peut-être pollué » qui représente moins de 1% des fichiers analysés. Parmi les fichiers classés dans cette catégorie par la métrique, 78% ont été jugés sains

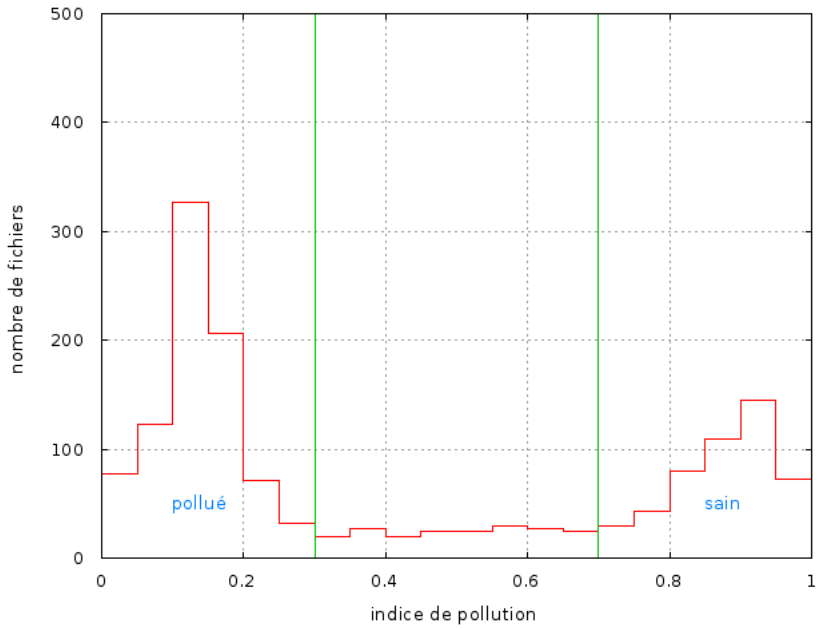

FIGURE 4.2 – Distribution du nombre de fichiers en fonction de l'indice de pollution obtenu

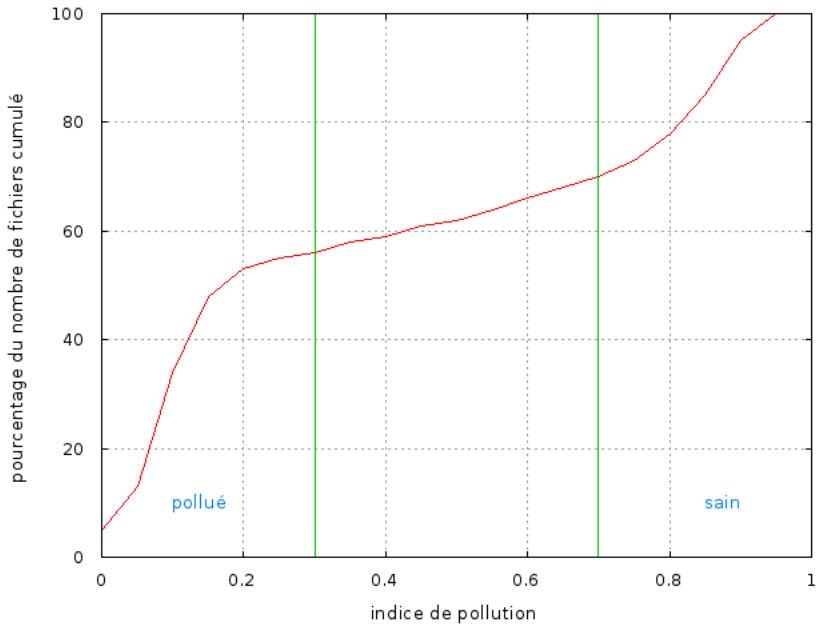

FIGURE 4.3 – Distribution cumulative du nombre de fichiers en fonction de l'indice de pollution obtenu

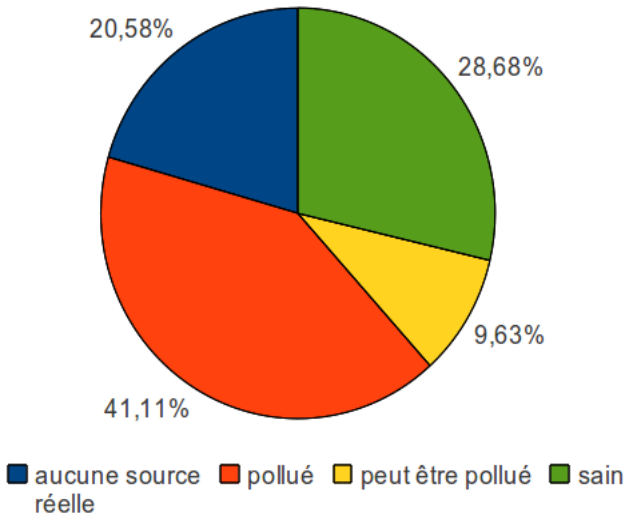

FIGURE 4.4 – Quantification de la pollution des contenus dans le réseau P2P KAD

% de faux positifs	% de faux négatifs
3.78	0.88

TABLE 4.1 – Taux d'erreur de la métrique de détection de la falsification d'indexation

par les expert et 12% pollués. L'avis des experts nous permet également de calculer le taux d'erreur de notre métrique. Ainsi, les faux positifs sont les fichiers considérés comme pollués par la métrique et sains par les experts. A l'inverse, les faux négatifs sont les fichiers considérés comme sains par la métrique et pollués par les experts. Le tableau 4.1 donne les valeurs des taux d'erreur. Au regard du taux d'erreur très faible constaté, la métrique que nous proposons détecte très bien la pollution par falsification d'indexation. La proportion de fichiers sains est en revanche sous évaluée car une partie d'entre eux est considérée comme peut être-pollué et un faible taux de faux positifs subsiste.

Nous sommes cependant capables d'identifier les rares cas de faux positifs. Il s'agit de version localisée de films dont le titre a été traduit. Ainsi, il peut arriver qu'une partie des utilisateurs nomme le fichier par le titre original du film alors qu'une autre partie utilise la version localisée du titre. Le tableau 4.2 illustre ce problème pour deux fichiers ayant mal été identifiés par notre méthode. Du point de vue de la métrique, les différents noms ne semblent pas liés car les mots-clés sont complètement différents, seule une connaissance sémantique des mots traduits ou des différents noms existant pour un même film permet de considérer que le contenu est sain, ce qu'ont bien identifié les experts.

Nous avons également regardé comment la pollution affectait chacun des 100 contenus étudiés au travers des 20 fichiers considérés pour chacun d'eux. Il apparaît que tous sont affectés avec entre 5 et la totalité des fichiers pollués par la falsification d'indexation. Ainsi, le contenu le moins pollué est « the big bang theory » avec 5 des 20 fichiers considérés pollués alors que l'ensemble des 20 fichiers sélectionnés pour « avatar » étaient pollués. La pollution n'affecte pas uniquement des contenus soumis aux droits d'auteur puisque le mot-clé « ubuntu » désignant

Nom de fichier choisi (X)	Nom de fichier majoritaire des sources (Y)
`Il Cigno Nero Sub Ita.avi`	Black.Swan.2010.DVDSCR.XviD-TiMKY.avi
`(DivX ITA) 2012. 2009.avi`	Segnali.Dal.Futuro.2009.iTALiAN.LD.DVDRip.XviD-SiLENT.CD1.avi

TABLE 4.2 – Exemple de faux positifs lors de la détection de la pollution

	% de fichiers pédophiles	% de fichiers pornographiques
pour tous fichiers	3,6	21,1
pour les fichiers pollués	8,8	55,7

TABLE 4.3 – Pourcentage des fichiers supervisés potentiellement pédophile ou pornographique

une distribution linux sous licence GNU GPL est également largement pollué avec 15 fichiers pollués parmi les 20.

Pour terminer, nous avons évalué la proportion de fichiers affectés par le mélange d'indexation dont au moins un des titres proposés indique un contenu pédophile ou pornographique. Nous avons recherché pour cela les mots-clés pédophiles explicites, et des mots-clés à caractère pornographiques utilisés par le logiciel ProCon [3] pour réaliser un contrôle parental. Les résultats sont présentés dans le tableau 4.3. Il apparaît que la majorité des fichiers pollués par la falsification d'indexation sont potentiellement des fichiers pornographiques. Plus grâve, 8.8% des fichiers pollués peuvent contenir au final un contenu pédophile ce qui, étant donné la popularité des fichiers étudiés, implique une diffusion importante de ces contenus illégaux à l'insu des utilisateurs et peut générer de nombreux faux positifs en cas de supervision inadaptée.

4.3 Modification du mécanisme d'indexation

Nous avons montré que notre méthode de détection de la pollution par falsification d'indexation était très fiable. Elle pourrait ainsi constituer un mécanisme de défense, ou au moins de prévention en indiquant à un utilisateur s'apprêtant à télécharger un fichier si celui-ci est suspecté de pollution. Cependant la structure actuelle du mécanisme d'indexation de KAD ne permet pas d'obtenir les informations nécessaires à l'application de notre métrique avant de télécharger un fichier. Cette détection tardive est dommageable car l'utilisateur peut être amené à être supervisé accédant à un fichier non désiré. De plus, si l'on considère le cas de petits fichiers (<1Mo), le téléchargement peut être très rapide et se terminer avant même la détection de la pollution. Notre méthode est également très coûteuse d'un point de vue réseau puisque des centaines de connexions TCP doivent être initiées afin de récupérer les noms de fichiers des sources, et ce, sans réelle intention de téléchargement dans le cas où le contenu est pollué.

Il serait cependant facile de rendre cette détection possible au niveau de la DHT. Il suffit pour cela de modifier très légèrement le protocole d'indexation de KAD en ajoutant un tag contenant le nom de fichier lorsqu'un pair se publie comme source de celui-ci. Les pairs chargés de l'indexation des sources du fichier reçoivent alors les noms de fichier contradictoires et

3. `https://addons.mozilla.org/fr/firefox/addon/procon-latte/`

peuvent appliquer notre métrique pour détecter la pollution. Dès lors, un simple message UDP serait suffisant pour obtenir l'indice de pollution d'un fichier en contactant les pairs proches de l'identifiant du fichier sur la DHT.

Cette solution a cependant une limite immportant du fait d'un autre problème de sécurité majeur affectant les DHT : l'insertion localisée de pairs pouvant prendre le contrôle des références indexées. En effet, si des pairs malveillants sont insérés à proximité de l'identifiant du fichier dont l'indexation est mélangée, ceux-ci peuvent facilement mentir sur la valeur réelle de l'indice de pollution afin de la masquer et rendre cette protection caduque. Cette solution doit donc être couplée à une autre empêchant l'insertion localisée de pairs dans le réseau [CCF10].

Chapitre 5

Conclusion et travaux à venir

Dans le cadre du projet GIS 3SGS ACDAP2P, nous étudions la possibilité d'utiliser une approche collaborative pour la détection d'attaques sur les réseaux pair à pair. Dans ce contexte, le premier travail a consisté à effectuer l'état de l'art des réseaux pair à pair et de leur sécurité en terme de failles et solutions collaboratives pour la détection. Le présent délivrable a présenté nos premières contributions allant dans ce sens à savoir l'identification des vulnérabilités de la DHT de BitTorrent ainsi que la détection des comportements malveillants dans KAD. Chacune de ces deux contributions a été validée par une publication.

Nous avons identifié dans cette section une nouvelle forme de pollution affectant largement le réseau P2P KAD, appelée falsification d'indexation. Celle-ci est très néfaste car elle amène un utilisateur à télécharger un contenu indésirable et potentiellement dangereux. L'étude de 2000 fichiers parmi les plus populaires de KAD a montré que plus de 41% d'entre eux sont victimes de cette forme de pollution. Cette quantification de la pollution est la première réalisée depuis l'étude menée sur Overnet en 2006 [LNR06] et montre que la pollution des réseaux P2P reste un problème majeur aujourd'hui, la pollution des contenus étant encore plus présente et néfaste qu'alors. Nous avons proposé et validé une métrique basée sur le coefficient de similarité de Dice qui est capable de détecter le mélange d'indexation avec une grande fiabilité. Cette détection est cependant réalisée tardivement lors du début du téléchargement mais pourrait faire partie intégrante du mécanisme d'indexation au prix d'une modification mineure du protocole de KAD. Cette solution contre la pollution est cependant vulnérable à un problème plus large affectant les données indexées dans la DHT à savoir, l'insertion ciblée de pairs malveillants. Nous allons étudier plus précisément ce problème et proposer une solution dans la suite de ce manuscrit.

Le prochain délivrable porte sur la proposition d'une solution collaborative permettant de détecter les comportements malveillants dans les réseaux P2P. Etant donné le nouveau contexte du réseau KAD, notre étude ne portera pas sur les attaques internes comme prévu initialement car celles-ci sont actuellement limitées, mais sur les attaques externes engendrant l'immense pollution constatée quotidiennement par les utilisateurs du réseau. Nous proposerons ainsi, dans le prochain délivrable, une métrique capable de détecter précisément les contenus pollués puis nous terminerons le projet en proposant une approche collaborative basée sur cette métrique et capable de limiter la diffusion de la pollution.

Annexe A

Liste de 100 contenus populaires téléchargés en 2010

inception
iron man 2
2010
xxx
french
avatar
dvdrip
despicable me
porn
clash of the titans
toy story 3
glee
salt
twilight eclipse
dexter
apprentice sorcerer
axxo
robin hood
prince of persia
windows 7
greek get him
predators
airbender last
shutter island
knight and day
expendables
takers
dinner for schmucks
unstoppable
eli book
grown ups
true blood
alice in wonderland
movies
shrek forever after
supernatural
hindi
house
devil
step up 3d
megamind
harry potter and the deathly hallows
skyline
green zone
naughty america
eminem
lost
town
date night
wolfman
smallville
last song
torrents
dragon how to train your
fringe
dear john
red
social network
weeds
noir
pc games
vampire diaries

2012
twilight
cop out
tamil
city 2 sex and the
remember me
walking dead
eclipse
due date
fxg
grown ups
entourage
sherlock holmes
how i met your mother
sex
microsoft office 2010
spartacus
pacific
karate kid
other guys the
call of duty black ops
chuck
ita
resident evil
wii
hot tub time machine
ubuntu
nero
ncis
theory big bang
indiana jones
tron 2
lady gaga
raiponce
photoshop
black swan
tourist the

Bibliographie

[CCF10] Thibault Cholez, Isabelle Chrisment, and Olivier Festor. Efficient DHT attack mitigation through peers' ID distribution. In *Seventh International Workshop on Hot Topics in Peer-to-Peer Systems - HotP2P 2010*, Atlanta États-Unis, 04 2010. IEEE International Parallel & Distributed Processing Symposium.

[LKXR05] Jian Liang, Rakesh Kumar, Yonjian Xi, and Keith W Ross. Pollution in p2p file sharing systems. In *IN IEEE INFOCOM*, pages 1174–1185, 2005.

[LMSW10] Thomas Locher, David Mysicka, Stefan Schmid, and Roger Wattenhofer. Poisoning the Kad Network. In *11th International Conference on Distributed Computing and Networking (ICDCN), Kolkata, India*, January 2010.

[LNR06] Jian Liang, Naoum Naoumov, and Keith W. Ross. The index poisoning attack in p2p file sharing systems. In *INFOCOM*. IEEE Computer Society, IEEE, 2006.

[MCD+11] Guillaume Montassier, Thibault Cholez, Guillaume Doyen, Rida Khatoun, Isabelle Chrisment, and Olivier Festor. Content Pollution Quantification in Large P2P networks : a Measurement Study on KAD. In *11th IEEE International Conference on Peer-to-Peer Computing (IEEE P2P'11)*, pages 30–33, Kyoto, Japon, August 2011. IEEE Communications Society. Projet GIS 3SGS ACDAP2P.

[MS99] Christopher D. Manning and Hinrich Schütze. *Foundations of statistical natural language processing*. MIT Press, Cambridge, MA, USA, 1999.

[Tve77] Amos Tversky. Features of similarity. In *Psychological Review*, volume 84, pages 327–352, 1977.